SOURCES AND USES OF THE DATA

INTRODUCTION

The initial sample design called for a list of firms stratified by firm size. The sample was to be taken from a state registry or from registries maintained at oblast or rayon level. Because these data would represent only registered firms, a household sample was desired to supplement the government sources in order to measure the informal or unregistered portion of industry in the Ukraine. The term "industry" is used in this report to represent any enterprise selling or bartering goods and/or services in the marketplace. The reliance on the government registry was due to the need to stratify the sample by firm size. Stratification is necessary to obtain a representative number of the largest firms. Such firms would occur with a very low frequency in a purely random sample. Adequate representation of firms in each size class would be necessary to characterize the differences between the large and the small firms. Five thousand interviews were planned: one thousand in each of five different size classes: zero employees (owner only), 1-10 employees, 11-50 employees, 51-250 employees, and over 250 employees)

The household sample would be completely random. In the interest of costs, some clustering around randomly selected starting points would be permitted, but no cluster was to be larger than 10 households (0.025% of the total sample). The sample would be drawn in proportion to the population in each region and from the majority of the postal districts. The population in the household would be used to develop projection factors based upon the ratio of the total population to the number of householders. The household survey would be necessary to locate those entrepreneurs too new to the market or too small to be included in a government registry or those firms of any size that were deliberately not registering their activity with the government. Measurement of unregistered or informal market activity was a major goal of the study. A survey of four thousand households was planned for this part of the study.

Due to the interest in small firm characteristics and because the smallest firms are found in street trade, a small street sample was planned. This would permit the ability to characterize the operating characteristics in small firm sales and services that could be observed along the street or in the marketplace or similar commercial areas. Because this sample was not tied to any official registry, or in proportion to any specific population measure, the results would not be used for any national or regional projected measures. The data would be used to enrich our knowledge of the operating characteristics of the smallest enterprises.

THE REGISTRY SAMPLE

After interviews with several government offices it was determined that the best available single list of firms in the Ukraine was that of the Committee of Statistics. Since these were firms that were registered with some level of government, this list is referred to as the "register". Tax authorities in the individual oblasts were reported to have registered approximately 800,000 firms. Because larger firms might have separate establishments within or across one or more

oblasts, the actual total for the Ukraine might not be the simple sum of the regional measures. As a result it was decided to use a central register such as that of the Committee of Statistics.

The marketing personnel at the Committee of Statistics reported that their file contained 600,000 records of individual firms. While this was less than the 800,000 previously reported, this difference could be the combination of establishments into their parent enterprises. To ascertain the viability of the registry a summary by firm size was requested before entering into any agreement.

The results of this advance summary called for a change in sample design. Instead of the 600,000 firms originally cited, only 203,168 firms could be sorted according to employment size. The results of this advance computer run were:

Employment Size of Firm	Number of Firms
Zero	10,288
1-10	110,370
11-50	44,098
51-250	27,561
Over 250	10.851
Total	203,168

These numbers would be useless for national projections. Our knowledge of market economies elsewhere in Eastern Europe and in the West indicated that there should be some millions of firms with 10 or fewer employees. While we could draw a sample of firms for interviews from the registry, we could not use the registry count of total firms to project national totals. We would have to use the household sample for estimates of employment in smaller firms and for identification of entrepreneurs with smaller enterprises. The registry sample could still be used for firms of employment size over 250. The decision on whether to use the registry sample for the 51-250 employee size group would depend upon the results of the household survey. The relationships between the different size groups form a common pattern in virtually every economy. This pattern would help to determine whether the household or the registry sample would be used to project the 51-250 employment sample size group to national totals.

The registry sample was important for much more than projection of national totals. The sample could be stratified by firm size. If we relied on a purely random sample, we would get very few firms in the largest size group. If the count of 10,851 largest firms was correct, a random sample of 5,000 firms from the registry would yield only 267 firms to be interviewed in that size class. This would give inadequate coverage of industry and region. The goal of the sample was to achieve 1,000 interviews in each of the size classes listed above. Because names and addresses would be supplied, the interviewers could locate specific firms and ask for a company spokesperson for the interview. The many questions about employment, economic outlook, history of ownership, and company performance could not be expected to be obtained from an employee in a household interview. A senior executive in the larger firm could be expected to be knowledgeable about these areas, however.

The inaccuracy observed in the registry data is to be expected. The State Committee for Statistics in the Ukraine has had only a few years to establish a database of market-oriented companies. Statistical experts familiar with problems of firm measurement in other countries can state unequivocally that such a database is difficult and expensive to maintain. Even the best public and private databases in the West have noticeable errors at any given time. As an example of this difficulty, a newly formed cohort of firms in the United States. has a half life expectancy of 38 months. The rate of discontinuance is estimated at 20% per year, compounded. Such rapid change is characteristic of dynamic markets as industry responds to new technologies, new methods of distribution, and changing tastes and preferences. To allow for obsolete entries in the registry database the sample drawn from the registry had to be much larger than our target of 1,000 for each employment size group. The Zero and the 1-10 group were combined for a target of 2,000 completed interviews. Table D-1 illustrates the problem and is drawn from the information in Tables 2.2.4 and 4.1.1 of Appendix A, the *Methodological Report.*

Table D-1.   Experience with the Registry Sample

	Size of Firm				
Item	**Zero-10 Employees**	**11-50 Employees**	**51-250 Employees**	**Over 250 Employees**	**Totals**
A. Final Registry Count	98,455	42,188	25,619	10,269	176,521
B. Delivered Sample	5,593	2,802	2,791	2,814	14,000
C. Assigned to Interviewers	4,766	1,862	1,779	1,709	10,116
D. Non-Existent Enterprises	2,610	560	336	164	3,670
E. Completed Interviews	1,187	895	934	961	3,977
F. Targeted Number	2,000	1,000	1,000	1,000	5,000
G. = (C – D)/C = Percent Good in Sample	45.2	69.9	81.1	90.4	67.0

Less than half of the records in the smallest size group represented existing companies. The three larger firm size groups were increasingly better. For the whole sample, however only two out of three firms in the registry could be found. Research in Western economies has measured the increased stability of firms as size increases. The higher proportion of firms found in the larger size firm size categories from the registry show this to be true of the Ukraine as well.

The registry count is noticeably lower in these final sample searches than the 203,168. When the computer search asked for industry as well as firm size, the number of "hits" was reduced to 176,251, a reduction of over 13 percent. The smallest group lost the most records and the largest group the fewest.

THE HOUSEHOLD SAMPLE

Based upon an original plan for a sample of 10,000 interviews and with 5,000 planned for the registry sample, the remainder of 5,000 interviews was considered for household and street interviews. Because the street interviews could not be used for national or regional projections 4,000 interviews were scheduled for the household survey and only 1,000 for the street survey. The household survey was divided proportionally to the population in each of the 24 oblasts, Kiev City, and the autonomous region of Crimea an grouped into five regions as described in Table 2.2.9 of Appendix A *Methodological Report*. The population measure chosen was population age 15 and over, a measure of potential entrants to the workforce.

The household sample was necessary not only for verification of the registry sample in the smaller size groups, but also to locate nascent entrepreneurs and employment in firms not entered in the registry. Altogether 4,002 households were visited with interviews of 9,789 persons age 15 or over. Some of the interviews were for an absentee household member through a fellow householder as spokesperson. Of these 9,789 potential workers, 3904 reported that they were employed in an industry supplying goods or services in markets. An additional 646 household member reported ownership of a firm as entrepreneurs.

The household sample was centered in some 200 postal districts. Random geographic locations were chosen for starting points and a cluster sample not exceeding 10 households was taken around that specific geographic point. By limiting cluster size to a maximum of 10 households, adequate geographic dispersion of the 4,000 households was assured.   Such limited clustering was undertaken to hold down field costs. National and regional projections were made in proportion to the ratio of the population to the sample size. Regional projections were made only from the household sample, while national projections were made from a combination of household and registry sample as described in the next section.

THE STREET SAMPLE

It was anticipated that the majority of small firms would be in the Wholesale and Retail Trade industry. The registry sample did not cover very small firms well. The household sample would interview mostly employees rather than managers. To get manager perspective to the problems and outlook for small business a small street sample was designed. Because the sample was not tied to population or to the registry and because an adequate random sample would be expensive, an area sample was designed. The interviewers would start at a hundred or more random locations, enumerate up to 30 businesses at each location and interview up to ten businesses immediately adjacent to the randomly selected location. Location selection was limited to areas of commercial activity.

The street sample is not analyzed in this report, since this report focuses on analysis of businesses across all industry types, using a sample that can serve as the base for national and regional projections.

PROJECTION OF SAMPLE RESUILTS TO NATIONAL AND REGIONAL TOTALS

The projection of sample results to regional totals is covered in Appendix C. Because industry totals were desired, only the household sample could be used. Future work should incorporate computer sorting of records from the registry file to correct the projected totals in the largest size class. The registry file has not been analyzed for regional observations by industry and size class and was unavailable for this purpose. For the present, the regional projections from the household sample are reasonably accurate for all except the over 250 employee size class.

The reasons for the failure of the household sample to adequately measure employment in the largest size class could be due to several factors. It is possible that household member employees misjudged the size of their employing firm. This could be due to the fact that they worked in an establishment smaller than the entire firm. It could be that the number 250 seemed large from their viewpoint. It could also be that the interviewer was not prepared to pursue the question of firm size with care.

The projected national totals given in the main body of the report in Tables 1.1, 2.1, and 2.2 are based upon projections from both the household and the registry sample. As can be seen from table D-1, the registry records were inadequate for the smallest size groups. The registry was marginally acceptable for the 50-251 size group.

Projection for the zero employee size group was taken from the household sample of 4,000 where data are available for 636 entrepreneurs. These households contained 9,789 persons age 15 or over. The national projection for zero employee firms is 636 times the ratio of national population age 15 and over of 40,809,592 to the number of persons in the 4,000 households age 15 and over of 9,789. The quotient of 4,169 is the projection factor. Each entrepreneur located in a household became the basis for 4,169 firms in the nation. This number could be refined by reducing the number observed in households to 593 rather than 636. This is because 492 of the entrepreneurs reported no full-time employees, while 143 reported one or more employees. An additional 101 zero employee firms were located from the registry sample. Tables 1.1, 2.1, and 2.2 in the main body of the report are based upon 636 zero employee firms. The product of 636 and 4,169 is 2,651,435 as reported. If we use the 593 as the basis, the product of 593 and 4,169 is 2,472,217 which might be an improvement over the previously reported figure, but is well within normal sampling error.

Projection for the 1-5 employee size group is based upon 124 employees in the sample. The projection factor remains 4,169 and the estimated employment is 516,947. Given the low occurrence and low discovery in the registry sample, there is no basis to offer an alternative to this number. The number of firms was not reported in the main body of the report for this size group. Based on the sample taken from the registry the mean number of employees in the 1-5 size group is 3.37. This number is "robust" as it is taken from 626 interviews with company managers. 516,947 divided by 3.37 gives 153,947 firms which is very close to the 148,976 reported in table 1.1 of the main report.

Projection for the 6-10 employee size group is based upon interviews of 204 household members. Projection by the factor of 4,169 gives an estimated total employment in the 6-10 employee size group of 850,476, the same as the 850,460 in table 1.1, allowing for rounding

error. The mean number of employees in firms of this size group using 524 interviews of company managers is 8.14 employees. 850,476 divided by 8.14 gives 104,481 firms in the 6-10 employee size group. This agrees with the 104,608 given in table 1.1 within normal rounding error.

Projection for the 11-50 employee size group is based upon 765 household interviews. The projection factor remains 4,169. The projected number of employees in the 11-50 employee size group is 3,189,285. This is within rounding error of the 3,189,226 employees in this size group reported in table 1.1 in the main body of this report. From 981 interviews with company officers or spokespersons the mean number of employees in the 11-50 employee size group is 25.95. Division of total employment by the average employee number gives 122,901 firms. This is within rounding error of the 123.757 originally reported in table 1.1.

Projection of employment in firms in the 51-250 employment size group could be made either from the household sample or from the registry sample. The household survey interviewed 1009 members employed in firms of 51-250 employees. Multiplied by the projection factor of 4,169 produces employment of 4,206,521 in this employment size group of firms. This compares closely with the 4,206,444 reported in Table 1.1. The average number of employees in the 51-250 employee size group is 125.05. Dividing the number of employees by 125.05 gives an estimate of 33,638 firms in the 51-250 employee size group. This compares closely with the 33,169 reported in Table 1.1. We note that over 80 percent of the records from the registry sample were reliable, so a check projection can be made from the registry to see how the estimates compare.

The registry sample reported 929 firms with employment between 51 and 250 persons. The average employment in this employment size group was 125.05. If we assume that the 28,338 records originally reported represent a measure of all valid records whether they could be sorted or not, we can develop an alternate estimate. 929 firms times the ratio of 28,338 to 929 times the average employment of 125.05 gives an estimate of 3,543,667 employees in the 51-250 employment size group. This is well below the 4,206,444 estimated from households. The calculation shows that registry coverage is approaching household estimates as firm size increases, but cannot be recommended for any but the largest size employment group

The projection of employment in firms of 250 or more employees is taken from the registry sample. While the registry originally stated that there were 10,851 firms with employment of 250 or more, the final sorting revealed only 10,269. If this were the entire number of firms, our use of 10,851 would represent an overestimate of 5.7 percent. But there is always the possibility that there were some large firms in the over 600,000 firms in the registry that could not be sorted by firm size. To put a lower estimate on the number of firms with 250 or more employees, we will use the lower figure 10,269. We can further reduce this by considering that 9.6 percent of the listed firms did not exist. This would lower the count to 9,283. Further examination of the registry sample revealed that 77 records or 4.5 percent of the sample were noncommercial. Removing this percentage gives a lower limit to the number of firms of 8,865. This is 18.3 percent lower than the 10,851 firms used in the projections in table 1.1, 2.1, and 2.2. But this number is far above the 5,011 large firms calculated from the household survey.

We are confident of the lower bound of 8,865 firms because the registry data was tested in the field. Names and addresses were verified. The loss of data was only the 9.6 and 4.5 percents discussed above. This is a lower bound because there may be many more large firms among the 600,000 records that could not be sorted by firm size. On this basis the household survey estimates for firms of over 250 employees and the employment these firms represent is rejected. The reader is left to choose between 10,851, 10,269, 9,283, and 8,865 as the number of very large firms. The employment associated with these numbers of firms at the rate of 905.22 employees per firm is 9,822,542 (the number in table 1.1), 9,295,704, 8,403,157, and 8,024,775. These are all well above the 4,535,789 that is projected from the household sample. Our choice for the largest of these numbers is due to the uncertainty associated with the 600,000 records that could not be sorted by firm size. The result is a conservative estimate of the relative importance of SMEs.